THE POLLS—REVIEW
WAS 1996 A WORSE YEAR FOR POLLS THAN 1948?

WARREN J. MITOFSKY

After the ballots were counted in the 1996 presidential election, there were no pictures of a victorious Bob Dole, gleefully holding a newspaper with the erroneous headline, "Clinton Defeats Dole." We saw no postelection speeches in which President-elect Dole ripped into the liberal media polls that had declared him prematurely dead. House Republicans held no hearings to get to the bottom of the polls' failure. There were no press conferences in which pollsters expressed regret and confusion about what could have led them to think that Bill Clinton would win. And a humbled Clinton did not opine that the lead he had in preelection polls must have lulled his supporters to sleep on election day.

We saw none of these things because, in concert with the estimates from all of the media preelection polls, Clinton won the 1996 presidential election. Unlike the 1992 British preelection polls, almost all of which erroneously foretold a Labour victory, or the 1990 Nicaraguan polls, many of which picked Daniel Ortega to defeat Violetta Chamoro, or the 1980 U.S. polls that predicted an uncertain Reagan lead and not a landslide, or the infamous 1948 polls that predicted a Dewey victory over Truman, the 1996 polls were unanimously correct in predicting that Clinton would win by a safe margin. Unlike those storied exemplars of polling frailty, earlier occasions for embarrassment and consternation, polling in the 1996 election could be judged a clear-cut success.

It came as a considerable shock, then, when Everett Carll Ladd, Jr.—director of the Roper Center at the University of Connecticut and a prominent figure in the field of public opinion research—declared in an article published in the *Chronicle of Higher Education* (1996a), and excerpted in the *Wall Street Journal* (1996b), that "election polling had a terrible year in 1996. Indeed, its overall performance was so flawed that the entire enterprise should be reviewed by a blue-ribbon panel of experts." A flurry of "copycat" negative media coverage followed Ladd's extraordinary statement, with articles appearing in *U.S. News and World Report,* the

WARREN J. MITOFSKY is president of Mitofsky International.

Public Opinion Quarterly Volume 62:230–249 0033-362X/98/6202-0006$02.50

New York Times Sunday Magazine, and a dozen or more periodicals and radio and television broadcasts.

Ladd's attack on the polls was a miscellaneous collection of charges. Most prominently, while acknowledging that Clinton did actually win the 1996 election, he said that the polls had overestimated the president's margin of victory. He remarked that, by comparison, the 1948 polls seemed ''closer to a triumph than a disaster.'' He said that Clinton's large lead in the polls throughout the campaign was illusory, but that its publication had dampened interest in the election and turnout on election day. He decried the number of preelection polls, whose findings ''bombarded'' the electorate throughout the campaign. He assailed not only preelection but also exit polls, charging that they had overestimated the Democratic share of the vote ''in recent years.'' He opined that Republicans and conservatives were less likely to agree to respond to polls (perceiving them to be tools of the liberal media), leading to nonresponse bias in preelection poll estimates.

Ladd's wide-ranging critique contained both testable charges and speculative complaints. Characteristic of the latter criticisms was his claim that the polls had overestimated Clinton's lead during the campaign and had thereby dampened interest in the election. His conjecture that the polls had suffered from nonresponse bias due to noncooperation by conservatives was similarly unverifiable. The impact, if any, of poll ''bombardment'' on prospective voters would have been difficult to assess during the campaign and was impossible to gauge retrospectively. To respond to any of these assertions, one would have to match speculation with surmise.

In addition, Ladd's testable claims themselves were poorly specified. His complaints about inaccuracy and Democratic bias in exit polls referred to problems in ''recent years'' and offered no data. And the marquee attack in Ladd's offensive—that 1996 poll accuracy was worse than in 1948 or any other year—was simply unsupported. His *Chronicle of Higher Education* article contained no 1996 data whatsoever. The excerpt of that article in the *Wall Street Journal* featured a table, titled ''Polls Away from Reality,'' listing final 1996 poll projections for eight polling organizations and the share of the vote for Clinton, Dole, and Perot. Ladd's ''analysis'' consisted solely of two statements, that ''most of the leading national polling organizations made pre-election estimates that diverged sharply from the actual vote on November 5,'' and that ''of late, both pre-election surveys and exit polls on Election Day frequently have missed the mark by margins well in excess of the Gallup results in 1948.'' He presented no criteria for assessing ''divergence'' nor any measures to support this claim.

Because even his highlighted charges were imprecise and undocu-

mented, a reasoned assessment of Ladd's attack requires some reformulation and refinement to enable gathering appropriate evidence. The responsible analyst, in other words, has to do the difficult work that Ladd had spared himself in leveling his charges and then make the effort to judge the case once the terms are clear. The prospect of performing such double duty is never inviting. But because Ladd's unexpected broadside received much national media attention, the National Council on Public Polls (NCPP 1997) felt compelled to defend 1996 poll performance. Considering Ladd's leading charge, that final preelection polls had badly overestimated the extent of Clinton's victory and were farther off the mark than in 1948, the council had to decide on a way to measure poll performance and then do the computations to provide a historical comparison.

The council's work was made more difficult by the fact that—after more than 50 years of election polling—no standard metric for gauging poll accuracy had been adopted by the polling community. When challenged by Ladd, the NCPP had to agree anew on a method of assessment. The absence of a generally accepted standard provides an open field for charges and countercharges based on hyperbole and blurred distinctions. But it is clearly an undesirable state of affairs for scholars who wish to study the question of poll accuracy with fairness and precision. My analysis, harking back to the Social Science Research Council (SSRC) report on the polls of 1948, examines the question of how poll accuracy should be measured, weighing the pluses and minuses of various approaches. I then revisit the Ladd-NCPP exchange and offer an answer to the question, Was 1996 a worse year for the polls than 1948?

Measures of Poll Accuracy

The SSRC study (Mosteller et al. 1949) of the 1948 preelection polls considered eight methods for measuring polling error. It noted that each method has ''advantages and disadvantages.'' The SSRC committee's definitions are as follows.

Methods for Defining Election Polling Error

1. The difference in percentage points between the leading candidates's share of the *total vote* from a poll and from the actual vote.
2. The difference in percentage points between the leading candidate's share of the *major party* vote from a poll and from the actual vote. (Major parties are Democratic and Republican and are assumed to be the top two vote getters.)

3. The average (without regard to sign) of the percentage point deviation for each candidate between his/her estimate and the actual vote.
4. The average difference (without regard to sign) between a ratio for each candidate and the number one, where the ratio is defined as a candidate's estimate from a poll divided by the candidate's actual vote.
5. The difference between two differences, where the first difference is the estimate of the vote for the two leading candidates from a poll and the second difference is the election result for the same two candidates.
6. The maximum difference in percentage points between a party and the actual vote.
7. The chi-square to test the congruence of the estimated and actual vote distributions.
8. The difference between the predicted and actual electoral vote.

A significant problem for today's evaluation of polling accuracy, and not addressed by the SSRC because the problem did not arise in 1948, is how to handle the ''undecided'' vote in the polls. There is no undecided number in an election; it only exists in some, but not all, polls. If the undecided respondents are not allocated to a candidate, then the average error computed using methods 1, 3, 4, 6, and 7 will be exaggerated. For example, suppose that an election that was 55 percent Democratic and 45 percent Republican had a poll that showed 50 percent for the Democrat, 40 percent for the Republican and 10 percent undecided. If there was no allocation of the undecided, methods 1 and 3 would have errors of 5 percentage points. If the undecided were allocated proportionally and the error computed, methods 1 and 3 would only show an error of 1 percentage point. Furthermore, the error computed using any of methods 1, 3, 4, 6, and 7 will not be comparable for polls that report ''undecided'' and those that do their own allocation.

The alternatives for evaluating polls that include an undecided category are as follows: (1) allocate the undecided in proportion to the votes for candidates in a poll, (2) allocate the undecided evenly between the two major parties, (3) allocate all the undecided to the challenger, if there is an incumbent, or (4) use one of the methods that does not require that the undecided be allocated. A more complex allocation cannot be accomplished by an evaluator.

Crespi (1988, p. 22) claims that the pollsters he interviewed for his book thought that proportional allocation of the undecided was ''closest to the experience of most pollsters.'' Actually, the arithmetic can be done for all methods without allocating the undecided, but, as shown above,

it does not result in comparable measures. The overriding consideration concerns the comparability of the measure over several elections.

Comments on Methods for Measuring Accuracy

The SSRC chose the straightforward analysis and interpretation of method 1 for its evaluation of the 1948 preelection polls. But they also expressed concern about the use of method 1 in elections with significant minor party candidacies (in 1948, Strom Thurmond and Henry Wallace). Method 1, while simple and easily understood, is artificial unless the test in the election is whether the leading candidate gets 50 percent or more. If there are more than two candidates in an election or if a poll includes an undecided percentage, the number for the leading candidate alone is of little value in describing the status of an election. Witness a recent argument between the Eagleton and Zogby Polls following the 1997 New Jersey governor's election. The Eagleton Institute of Politics (1997) claimed that it was closest with a preelection estimate of 45 percent of the vote, but that figure is of little use unless we also know that Christie Todd Whitman's opponent, Jim McGreevey, had significantly less than 45 percent. The Eagleton claim of accuracy (based on method 1) does not address this question (although their press release did include the vote for other candidates).

The SSRC committee also liked method 2 and used it, too, in their analysis. Method 2 reduces the percentages so the top two candidates add to 100 percent. The unspoken advantage is that this method eliminates all other candidates and any undecided who may be in a poll. The measure, unlike method 1, means the same thing regardless of how many candidates participate in an election: at least one candidate will have 50 percent or more. While method 2 may appear to be similar to method 1, there is an important difference. When poll numbers are repercentaged so the two major party candidates add to 100 percent something very important happens: the undecided percentage is eliminated. The effect is that method 2 becomes identical to method 5 with the undecided allocated. If the undecided are not allocated before applying method 5, then the two measures are not equivalent.

Method 3 averages the percentage point deviation for each candidate between its estimate and the actual vote, without regard to sign. This approach, the SSRC committee thought, had "inherent drawbacks. By including many small parties which scarcely contribute to the total vote . . . the average deviation can be made very small even though major party predictions have large errors" (Mosteller et al. 1949, p. 56). They understated the problem. If all 22 parties on the ballot in 1996 had been included in a computation using this method, then the average error would have been close to zero. Even with a limited number of candidates, if the coef-

ficient of variation[1] is roughly constant for each candidate included, then the overall error will decline as the number of candidates increases. Clearly, there needs to be a limitation on the number of candidates. This method, if used with ''discretion'' about which parties to include, the SSRC report said, could be useful (p. 56). It should be noted that neither the SSRC nor the NCPP, which also used this method in its evaluation of the 1996 preelection polls, defined what might be reasonable criteria for including third parties.

Method 3 has the virtue of evaluating candidates other than the top two. If the intent of a preelection poll is to report the standing of each candidate, this measure evaluates the average accuracy of more than two candidates. The criterion for including more than two candidates is arbitrary. Crespi (1988) included third-party candidates only if they received at least 15 percent of the vote in an election. In the twentieth century, his third-party criterion would have included Theodore Roosevelt (1912), Robert LaFollette (1924), and Ross Perot (1992). It would have eliminated Strom Thurmond and Henry Wallace in 1948, George Wallace in 1968, John Anderson in 1980, and Ross Perot in 1996, among others (Congressional Quarterly 1985). Method 3's other failings are (1) a lack of comparability between elections that have different numbers of meaningful candidates, and (2) like some other methods, it requires the analyst do something the pollster who created the poll was unwilling to do: namely, allocate the undecided voters among the candidates. If the undecided were not allocated, the measurements of error would not be comparable from poll to poll.

Method 4 computes the ratio of each candidate's estimate divided by the actual vote; error is the average deviation from one for each candidate. This approach tends to exaggerate small percentage point differences in minor party candidates. For example, a one-point error in a party with 50 percent of the vote results in a 2 percent error. A one-point error in a party with 5 percent of the vote produces a 20 percent error. If all parties' errors are averaged, the overall result exaggerates the total error. This is just the opposite of method 3, which minimizes the total error. Method 4 could be modified to include major party candidates only, but then its result would be comparable to methods 2 and 5 (with allocation of the undecided).

The only problem with method 5, according to the SSRC report, was the ''complexity'' of explaining it. Method 5 first computes the difference between the two leading candidates in the poll and the actual vote; the error is the difference between these differences.

1. The coefficient of variation is the standard error of an estimate (percentage) divided by the estimate (percentage). This is a more useful measure of variability than the standard error alone because coefficients are comparable from one candidate to another.

In a race involving only Democrats and Republicans, after allocation of the undecided, method 5 yields results exactly two times the result of methods 1 and 2. An advantage of method 5 is that it evaluates the statistic most often reported by the media when reporting preelection polls, which is the margin between the top two candidates. Ladd, in his *Wall Street Journal* article, bases his discussion on the margin between the top two candidates. More recently, the *Newark Star-Ledger,* in its report of the Eagleton Poll for the 1997 New Jersey governor's race stated in its lead paragraph that Whitman had ''a 9-point lead'' (Hassell 1997). In the fifth paragraph the story cited Whitman's 45 percent figure, which the Eagleton Institute of Politics (1997), in its postelection press release, said was the most important figure. The author's nonscientific review of final preelection poll stories in 1996 and 1997 showed that almost all poll stories made primary mention of the margin between the top two candidates. The fact that there were distant third-party candidates was mentioned in these stories, but without comment, presumably because those candidates appeared to have no chance of challenging the two leading candidates. The margin between the leading candidates also was the measure pollsters used when they wrote about polls (DiVall 1996; Newport 1997; Taylor 1997).

Method 5 rewards the effort of the pollsters who allocate undecided voters well and penalizes those who allocate poorly. If the pollster does not allocate at all then the margin reported by the pollster, presumably, is the best indication of the pollster's expectation about the election outcome. Method 5 does not force someone evaluating polls to make assumptions about the undecided that were not made by the pollster.[2] It should be noted that the results of methods 3 and 5 are identical for two-candidate races when the undecided are allocated. They differ when there is no allocation.

Methods 6, 7, and 8 were mentioned in the 1948 SSRC report but were not considered as viable options. Method 6 applies to the one party with the largest error, even if it is a minor party. Complexity was the reason for dropping method 7. The error in electoral votes projected and evaluated in method 8 is one step removed from evaluating national or state polls directly.

Crespi (1988), after he performed proportional allocation for the polls that included undecided voters in their base, evaluated methods 1, 3, and 6. He concluded that there was not much difference between them and used method 1 for his analysis, as did the SSRC committee for its report

2. I would like to acknowledge that I made the mistake of not allocating the undecided during the 15 years I directed the CBS half of the CBS/*New York Times* Poll. I now believe that it is unreasonable of a pollster to ask a reader or viewer of a final preelection poll to make an interpretation about how the undecided will vote. A poll is being reported so the public knows what to expect when the election takes place. Leaving the undecided in the base of the percentages reported does not serve the public expectation or the pollsters' claims about accuracy.

on the 1948 elections. Methods 3 and 6, according to Crespi's data, had the smaller coefficients of variation. Method 1, which he used throughout the book, had a slightly larger coefficient of variation than the methods he rejected for his analysis.

An analysis of recent British elections yields a mixed bag. Crewe (1997) prefers method 3. He calls it "the true test of a poll's accuracy" (p. 580). Nonetheless, the entire discussion in his article about the 1997 elections is about Labour's *lead,* the lead being the difference between the top two parties, which is evaluated directly by method 5. He offers computations of polling error for both method 3 and method 5. Robert M. Worcester endorses Crewe's preference for method 3.[3] He also added a few details: (1) polls are repercentaged so the *major parties* add to 100 percent;[4] (2) there are no decimal places in the poll's recomputed estimate of a party and there is one decimal in the election result.[5] He claims this method has been used for decades to evaluate British elections.[6]

Comparison of the Methods

Nine final preelection presidential polls from 1996 were evaluated using four of the eight methods. They include the eight polls listed in the table in the *Wall Street Journal* excerpt of Ladd's critique, plus the Politics Now/ICR poll. The nine polls were conducted by ABC News, CBS News/*New York Times,* Gallup/CNN/*USA Today,* Harris Poll, Hotline/Battleground, NBC/*Wall Street Journal,* Politics Now/ICR, Princeton Survey Research/Pew Research Center, Zogby Group/Reuters. Three of them, (Harris, Princeton Survey Research, and Zogby), reported vote for "other" candidates in addition to Clinton, Dole, and Perot. Four polls, (ABC, CBS, Hotline, and NBC), reported undecided voters in the base of their percentages. The final poll results, as reported by the polling organizations, are presented in table 1.

The accuracy of methods 1, 2, 3, and 5 were evaluated. Method 1 was used by the SSRC in 1948; method 2 is similar, but only deals with the major party candidates and also was used in the SSRC report; method 3

3. Robert M. Worcester, personal communication to author (E-mail), December 13, 1997.
4. Unlike U.S. elections, there are more than two major parties in British elections. In 1997, there were three major parties plus "others" in the evaluation.
5. The British evaluation, unlike the NCPP evaluation of U.S. presidential election polls, maintains the same number of significant digits for a poll that the poll had when it was reported by the media. The use of a constant evaluation method also makes it possible to readily compare polling performance over more than one election.
6. The evaluation of the performance of the polls for the 1992 British general election took on the same sense of urgency as the SSRC evaluation of the 1948 U.S. polls. The four final British polls showed Labour with a small lead, which suggested a coalition government. The Conservatives actually won by a small but comfortable margin. Butler and Kavanagh (1992) discuss these polls in chap. 7, "The Waterloo of the Polls."

Table 1. Final 1996 Preelection Presidential Polls

	Clinton (Democrat)	Dole (Republican)	Perot (Reform)	Other	Undecided
Election[a]	49.3	40.7	8.4	1.6	
ABC News	51	39	7		3
CBS News/*New York Times*	53	35	9		3
Gallup	52	41	7		
Harris	51	39	9	1	
Hotline/ Battleground	45	36	8		11
International Communication Research/ Politics Now	51	38	11		
NBC News/*Wall Street Journal*	49	37	9		5
Princeton Survey Research/ PEW Research Center	52	38	9	1	
Zogby/Reuters	49	41	8	2	

NOTE.—Data are percentages.
[a] Data are from Scammon and McGillivary 1997.

was used by NCPP and the British in their evaluation of election polls; method 5 deals with the statistic most often reported in the media, the margin between the leading candidates. The evaluation was done with and without allocation of the undecided voters. A few rules were observed in the calculations: the result of the presidential election was stated to within one decimal place. All poll numbers were whole percentages with no decimals, thereby maintaining the same number of significant digits as were published by the pollsters. Undecided voters were allocated in proportion to the vote for Clinton, Dole, and Perot, which was the allocation method least controversial and used by NCPP, the British evaluations, and Crespi (1988; vote for "others" in a poll was eliminated before allocation). Poll percentages after allocation were rounded to whole percentages before other calculations. A rank of 1 was assigned to the poll with the smallest error for a given method. If more than one poll had the same error they were each given the same rank.

A comparison of the effect on the rankings for a given method can be

seen in table 2. Allocation of the undecided voters changed the method 1 ranking after allocation by more than one place for four of the nine polls. Method 2 always includes allocation, and therefore the results are identical, excluding it from this comparison. The resulting ranking using method 3 is changed slightly by allocation, except for one poll. Method 5 shows the least variability in the rankings when the undecided are allocated. Only one of the nine polls shift their ranking by two or more places.[7]

A comparison of the resulting rankings for methods 1, 3, and 5 when there is no allocation shows considerable variability. Each method produces a different rank order. A comparison of all four methods when the undecided voters are allocated is much less variable, as one would expect. Method 1 varies slightly more than the other three methods, which is consistent with Crespi's finding.[8] Methods 2 and 5, as noted in the discussion of methods above, produce identical rankings after allocation.

Discussion

An inspection of the rankings in table 2 produced by the different methods shows that they are more consistent when the undecided are allocated. However, it is still open to question whether the evaluator should allocate the undecided if the pollster did not do it. If the goal of these preelection polls is to inform the public about the expected outcome of an election, then it seems that the responsibility for allocation should rest with the pollster. The public and the journalists have neither the information necessary for a sophisticated allocation nor the technical knowledge. If there is some goal for these polls other than forecasting the election, then the eight methods described above for evaluating the polls are not sufficient. A more appropriate approach would be to assess the conformity of the polling methods to good statistical and polling theory, which are the criteria to which all other polls are subjected. Assuming the goal is forecasting, it seems reasonable, when evaluating a poll, to take the numbers as reported by a pollster, without modifying them.

This conclusion was based on four points. First, five of the nine national polls did their own allocation of the undecided. Second, when evaluating the polls that did not allocate, there is no agreed upon method of allocation

7. An examination of the correlations among the ranking shows high consistency among the methods when the undecided are allocated. These correlations are all high, ranging from .72 to .94, except for the correlation between methods 2 and 5, which is 1.0. When the undecided are not allocated, the correlations are more variable and they are lower. They range from .10 to .75, with one exception. Again, the correlation between methods 2 and 5 is high; it is .94. I thank Dan M. Merkle for these computations.

8. The coefficients of variation for methods 1, 2, 3, and 5, respectively, are .70, .58, .53, and .64 for the 1996 final presidential election polls, when the undecided are allocated. Crespi's (1988) study had coefficients for methods 1 and 3, respectively, of .83 and.76.

Table 2. Rank Order of Accuracy of 1996 Preelection Presidential Polls, Based on Methods 1, 2, 3, and 5

	Undecided Not Allocated						Undecided Allocated					
	1	2	3	5	Average	Average Rank	1	2	3	5	Average	Average Rank
ABC News	3	4	5	4	4.0	5	8	4	6	4	5.5	7
CBS News/*New York Times*	8	9	9	9	8.8	9	9	9	9	9	9.0	9
Gallup	6	2	3	3	3.5	4	4	2	3	2	2.8	3
Harris	3	4	2	4	3.3	3	4	4	4	4	4.0	4
Hotline/Battleground	9	2	8	1	5.0	6	2	2	2	2	2.0	2
International Communications Research/ Politics Now	3	4	7	7	5.3	7	2	4	7	4	4.3	6
NBC News/*Wall Street Journal*	1	4	3	4	3.0	2	4	4	4	4	4.0	4
Princeton Survey Research/Pew Center	6	8	6	8	7.0	8	7	8	7	8	7.5	8
Zogby/Reuters	1	1	1	2	1.3	1	1	1	1	1	1.0	1

for an evaluator to use. Third, if there is a preferred method of allocation, the pollster did not choose to use it. Fourth, it seems reasonable to assume that pollsters report their ''best'' estimates of an election outcome to the public.

There is an effect on the measurement of accuracy of the polls when the undecided are allocated. The correlation of the errors for the nine polls using method 1 with and without allocation is only .35.[9] For method 3 it is .60. Only method 5 maintains a high correlation of .94. Should the evaluator do something that will change the assessment of accuracy that the pollster was unwilling or unable to do for him- or herself?

As to which method should be used to evaluate polling accuracy, that remains the choice of the evaluator. The arguments for and against each method are listed above and summarized here. If the goal is to forecast which candidate will win and by how much, then method 1 does not adequately evaluate an election outcome unless there are only two candidates. Limiting the analysis to one candidate, as method 1 does, gives no idea about the accuracy of the forecast; methods 2 and 3 evaluate the forecast of the winner indirectly and method 5 does it directly. Method 2 implicitly introduces proportional allocation, and therefore it too seems less preferable.

The best choice appears to be between methods 3 and 5. The chief argument made by proponents of method 3 is that it represents all ''significant'' candidates but leaves open how to define ''significant.'' Its opponents say method 3 artificially reduces the overall error when a third candidate is introduced, thereby making comparisons with two-candidate elections not meaningful. It should be noted, for example, that the introduction of Perot into the evaluation of the performance of the 1996 preelection polls reduced the measured polling error of method 3; the error on Perot's share of the vote is less than the overall per candidate error. Proponents of method 3 say it evaluates all candidates, which is true. It does *not,* however, provide a consistent method for evaluating a poll's forecast of the winning candidate in an election.

The choice then seems clearer. If one wants to compare elections over time it is necessary to use a method that is comparable for both two-candidate and multicandidate elections. Only method 5 meets that test.

Was 1996 a Worse Year than 1948?

Having reviewed the methods for judging poll accuracy suggested by the SSRC committee 50 years ago, I can now offer a reasoned judgment on

9. These are Spearman correlations, computed on the rank order of the polling errors. The correlations are very similar to the Pearson correlations using the errors themselves.

Table 3. Method 5 Errors in Presidential Polls, 1948 and 1996

	Truman-Dewey (%)	Method 5 Error (%)
1948:		
Election	4	
Gallup	−5	−9
Crossley	−5	−9
Roper	−15	−19
	Clinton-Dole (%)	**Method 5 Error (%)**
1996:		
Election	9	
Reuters/Zogby	8	−1
Hotline/Battleground	9	0
Gallup	11	2
ABC News	12	3
Harris	12	3
NBC News/*Wall Street Journal*	12	3
International Communications Research/Politics Now	13	4
Princeton Survey Research/PEW Center	14	5
CBS News/*New York Times*	18	9

the accuracy of the 1996 polls and assess the validity of Ladd's complaint, that the 1948 polls look better by comparison.

In 1948, George Gallup and Archibald Crossley had polls that were closer to the Truman-Dewey election outcome than the poll conducted by Elmo Roper. (See table 3.) Gallup and Crossley had Dewey winning by a 5 percentage point margin in a race Truman won by 4 points. This resulted in a 9-point error on the difference. Roper was farther off the mark. He had an error on the margin of 19 points. In 1996, eight of the polls had errors ranging between 0 and 5 points on the margin. The ninth, the CBS News/*New York Times* polls, had a 9-point error. The one poll with the largest error in 1996 was as far off as the best result in 1948. By this measure, the polls of 1996 were clearly better than the polls of 1948.

Polling closer to election day may have helped the polls of 1996. Gallup's 1948 national poll was concluded closer to the election than either

Crossley's or Roper's. Gallup stopped interviewing October 28, Crossley finished October 18, and Roper's poll was finished early in September. The 1948 election was November 2. In 1996, the eight national polls cited by Ladd completed interviewing during the closing days of the campaign. The earliest to stop was Hotline/Battleground on October 31, the Thursday before the election. Gallup polled through the night before the November 5 election, while most polls stopped on November 3.

National Council on Public Polls

The NCPP (1997), in an attempt to counter Ladd's criticism of the 1996 polls, published its evaluation of 47 final preelection presidential polls conducted between 1936 and 1996. The early polls (1936–60) in the NCPP analysis were only from Gallup. Harris polls were included from 1964 on; starting in the 1970s all other major national polls were included. For 1996, NCPP included eight polls cited by Ladd plus the Politics Now/ICR poll.

In its press release, NCPP (1997) ''refutes criticisms of the accuracy of 1996 national presidential polls'' (p. 1). It claims the average error ''was low relative to historical experience'' (p. 1) and within expected sampling error margins. Sheldon Gawiser, president of NCPP said, ''1996 should be remembered as one of the better years for the national polls'' (p. 2). The NCPP concluded, ''The average error in 1996 was only 1.7 percentage points. This compares to 2.5% between 1936 and 1996 and 1.9% since 1956. Eight of the nine [1996] polls had errors within the ±3% margin of error expected for samples of their size'' (p. 3.)

In its analysis, NCPP calculated polling error as ''the average [absolute] deviation between the final poll results and the election results for the top two or three candidates'' (p. 2). The third candidate was included in five of the 11 elections since 1956. The third candidates included in this analysis ranged from a low of 0.9 percent (McCarthy in 1976) to a high of 18.9 percent (Perot in 1992). There was one other wrinkle. Some polls allocated the undecided vote and others did not. In an effort to make all polls equal, NCPP allocated the undecided vote among the top two or three candidates in proportion to their estimated vote in the poll.

There was a debate within NCPP over which error concept to use. The members accepted method 3 (the average deviation between the poll and the election for each candidate) and rejected method 5 (the error on the margin) because it resulted in an average candidate error that was more than twice as large as the one used in their analysis.

Table 4 shows the average errors for each presidential election year between 1956 and 1996. The errors were computed using method 3, as NCPP did in its analysis, and method 5.

Table 4. Average Errors in Presidential Polls, 1948 and 1956–96

Year	Number of Polls	Number of Candidates	Method 3[a] Average Error (%)	Method 3[a] Rank	Method 5 Average Error (%)	Method 5 Rank
1996	9	3	1.7	5	3.6	8
1992	6	3	2.2	8	2.7	5
1988	5	2	1.5	3	2.8	6
1984	6	2	2.4	9	4.4	9
1980	4	3	3.0	11	6.1	11
1976	3	3	1.5	3	2.0	2
1972	3	2	2.0	7	2.6	4
1968	2	3	1.3	2	2.5	3
1964	2	2	2.7	10	5.3	10
1960	1	2	1.0	1	1.9	1
1956	1	2	1.8	6	3.5	7
Yearly average, 1956–96			1.9		3.4	
1948	3	3	4.9	12	12.9	12

SOURCE.—1996 from NCPP (1997) and publication; 1956–92 from NCPP; 1948 from Mosteller et al. 1949.

[a] Method 3 was used by NCPP in its analysis of the polls.

A few conclusions can be drawn from these data. The 1948 preelection polls stand out as the poorest performance of any preelection polls. Ladd's comparison of the 1996 polling performance to 1948 was without merit. The average error in the 1996 polls by either error measurement is much less than for 1948. Also, the error for each of the 1996 polls (except the CBS/*New York Times* poll) was less than their presumed sampling error.[10] The error on each of the 1948 polls exceeds what might have been the sampling error if those polls had been probability-based polls. To say that the polls of 1996 had estimates that "diverged sharply," as Ladd said, is wrong. They diverged modestly, and all but two overstated Clinton's lead over Dole. Only the CBS/*New York Times* Poll had an error approaching Gallup's 1948 error, and, unlike CBS and the *New York Times,* Gallup had the wrong winner.

However, the NCPP conclusion that 1996 was a banner year for the

10. The sampling error on the margin between the candidates is slightly less than twice the standard error on a single candidate. Polls that claim a "margin of error" of 3 percent (2 × standard error on one candidate) would likely have a margin of error on the diference of between 5 and 6 percent.

Table 5. Party "Bias" in Preelection Polls, 1956–96

	Number of Polls Favoring a Party		
Year	Democrats	Republicans	Neither
1996	7	0	2
1992	5	0	1
1988	1	4	0
1984	2	2	2
1980	4	0	0
1976	1	2	0
1972	1	2	0
1968	1	0	1
1964	2	0	0
1960	1	0	0
1956	0	1	0

polls also does not stand up to scrutiny. By its own error measurement (method 3), the average of the polls' errors in 1996 was the fifth best of the 11 presidential elections since 1956. By the author's preference for a measure of polling error (method 5), 1996 is eighth, somewhat worse than the NCPP's result. In either case, the 1996 performance is not "one of the smaller errors recorded," as NCPP President Sheldon Gawiser said in his organization's (1997) press release on polling accuracy.

Ladd also claimed that "election polls have frequently over-estimated the Democrats' share of the vote." The NCPP disagreed. It said, "Since 1956 errors favored the Democratic candidate in six elections and the Republicans in five. The size of the errors was almost equal."

Rather than examine the average error in each election year, as NCPP did, I examined the direction of the error in each final preelection poll since 1956. Averages have the potential for masking a potential bias, whereas individual poll results give a more direct picture. For this analysis, if a poll's margin between the two leading candidates varied by less than one percentage point from the election result, I said the poll favored neither party. (See table 5.)

The evidence shows that Ladd was correct about the direction of the polls' errors. More than twice as many polls overstated the Democratic candidate's share of the vote than overstated the Republican's share. Furthermore, the 25 Democratic-leaning polls had an average error on the margin of victory (method 5) of 4.4 percentage points, while the 11 Republican-leaning polls' error was only 3.3 percentage points. The Demo-

Table 6. State Polls, 1996

Error on Margin between Leading Candidates (%)	Number of Presidential Races	Number of Senate Races
10%+	4 (2)	12 (6)
7–9	16 (9)	14 (7)
4–6	18 (10)	27 (13)
1–3	47 (26)	45 (22)
<1	15 (8)	2 (1)
Total	100 (55)	100 (49)

NOTE.—Data are percentages, *N*s are given in parentheses.

cratic and Republican errors differ significantly and do not support the NCPP position that the parties' errors were about equal. Whether these differences are large enough to support Ladd's suggestion that the overreporting of Clinton's victory margin had a bearing on either participation or the outcome of the election is problematic. One would have to accept the bandwagon theory over the underdog theory in order to accept his notion, a discussion this article will not pursue.

1996 State Polls[11]

Ladd's (1996b) criticism of the 1996 polls included, by implication, the state polls as well as the national polls when he said, "Election polling had a terrible year in 1996." While he and his critics focused more attention on the national polls, the state polls are more numerous and appear more regularly in local news reports. The state polls, with very few exceptions, were done by different pollsters than the national polls. More than half of the state polls were done by Mason-Dixon, a firm that services news organizations nationwide. Mason-Dixon's performance was better, collectively, than those who did the other state polls. Of the 55 presidential state polls reported in the final Hotline (1996), 62 percent of them were within 3 points of the actual margin of victory. (See table 6.) The 49 state polls on senatorial races were not as good. Just under half were within 3 points of the final margin.

There were three instances where the presidential state polls had the

11. All poll results for this analysis come from Hotline (1996).

wrong winner as well as three errors in Senate polls. The incorrect presidential polls were all close races. Only one of the three Senate poll errors was close. The biggest error was by the *Detroit News* in the Michigan Senate race. They were off by 17 points on the margin. There were two others that were 14 points off, the *Omaha World-Herald* in the Nebraska Senate race and the *Greensburg (Pa.) Tribune-Review* in the Pennsylvania presidential race.

Ladd could hardly criticize the state polls for favoring the Democrats over the Republicans. The state polls were much more evenhanded than the national polls in the direction of their errors. Nineteen polls erred in favor of the Democrat, Clinton, while only 18 favored the Republican, Dole. The other 18 presidential polls were within 1 percentage point of the election.

The Exit Polls

Ladd did not spare the exit polls from his criticism. He said the networks offered premature reports on election night when their exit polling consortium, Voter News Service (VNS), incorrectly projected a Democratic win in New Hampshire. An ''especially egregious error,'' he called it. He also compared the performance of the exit polls in recent years to the 1948 preelection polls. He said they, too, had frequently missed the mark by larger margins than Gallup's error in 1948.

The networks used exit polls on election night for two purposes—projections and analysis. Exit-poll-based projections took place at poll closing time in contests that appeared to be clear-cut victories for a candidate. These projections have never cited exit poll estimates of the candidates' percentages. A VNS or a network analyst just named the winning candidate, which was then broadcast after the polls closed. The networks have never reported a margin of victory based on exit polls. Later on election night they did report estimates based on samples of actual vote returns, and these have almost always been within a few points of the final result. The analytical information used in cross-tabulation was weighted to the estimates produced from samples that used actual vote returns. As someone who did have access to the exit poll estimates, which were not publicly available, I can report that the 1996 exit polls were not in excess of Gallup's 9-point error in 1948.

Since the networks formed their exit poll pool in 1990 they have covered about 500 races. Over half were projected from exit poll results. The only incorrect projection by the pool was in the New Hampshire Senate race in 1996. This lone error was corrected by the networks on-air two and a half hours after the mistake was made.

Conclusion

Ladd's main point and most testable claim about the accuracy of the national polls holds no water: by any of the measures reviewed here, 1996 was not the best but was far from the worst year for the polls. The data also do not support a condemnation of 1996 state polls, nor do they support Ladd's claims about the accuracy of exit polls. The only charge that receives any support is Ladd's claim about overestimation of the Democratic vote share in polls. One measure, calculated for this article, bolsters this charge, while an NCPP analysis calls it into question. Overall, a modicum of scrutiny reveals that Ladd's impressionistic scorecard for the polls is seriously in error.

One can only speculate as to why Ladd chose to make demonstrably erroneous and unsupported claims. The professional polling community was understandably outraged in its response to Ladd's very public pronouncements. If he really meant to improve polling practice, one cannot imagine a less effective means of achieving the goal. Certainly, in view of his less-than-rigorous analysis, Ladd's call for a ''blue-ribbon'' commission to investigate poll performance cannot be taken seriously.

Issues facing the polling profession require dispassionate and careful study if appropriate improvements are to be found. In the first instance, we should be clear about how the quality of our work is to be measured. This article has examined a variety of rules and discussed their merits and drawbacks. The analysis provides a foundation for judging the quality of poll results. In 1998 and beyond, this framework may help to assess polling progress and problems. In the meanwhile, we can answer the question, Was 1996 a worse year for polls than 1948? No, it was much better.

References

Butler, David, and Dennis Kavanagh. 1992. *The British General Election of 1992.* London: Macmillan.

Congressional Quarterly. 1985. *Guide to U.S. Elections.* Washington, DC: Congressional Quarterly.

Crespi, Irving. 1988. *Pre-election Polling.* New York: Russell Sage Foundation.

Crewe, Ivor. 1997. ''The Opinion Polls: Confidence Restored?'' *Parliamentary Affairs* 4:569–85.

DiVall, Linda A. 1996 ''Keys to Expanding GOP Majorities.'' *Polling Report,* December 9, p. 1.

Eagleton Institute of Politics. 1997. ''Summary of the 1997 Gubernatorial Election Results Comparison with Pre-election Polls.'' Press release. Eagleton Institute of Politics, Rutgers University, New Brunswick, NJ.

Hassell, James. 1997. ''Latest Poll Shows Gains for Whitman.'' *Newark Star-Ledger,* November 2.

Hotline. 1996. ''#1'' and ''#11.'' *Hotline,* November 4.

Ladd, Everett C. 1996a. ''The Election Polls: An American Waterloo.'' *Chronicle of Higher Education,* November 22, p. A52.

———. 1996b. ''The Pollsters' Waterloo.'' *Wall Street Journal,* November 19, p. A22.
Mosteller, Frederick, Herbert Hyman, Philip J. McCarthy, Eli S. Marks, and David B. Truman. 1949. *The Pre-election Polls of 1948.* New York: Social Science Research Council.
National Council on Public Polls (NCPP). 1997. ''Polling Council Analysis Concludes Criticisms of 1996 Presidential Poll Accuracy Are Unfounded.'' Press release, February 13. National Council on Public Polls, Fairfield, CT.
Newport, Frank. 1997. ''Controversies in Pre-election Polling.'' Paper presented at the annual meeting of the American Association for Public Opinion Research, Norfolk, VA.
Scammon, Richard M., and Alice V. McGillivray. 1997. *America Votes 22.* Washington, DC: Congressional Quarterly.
Taylor, Humphrey. 1997. ''Why Most Polls Overestimated Clinton's Margin.'' *Public Perspective* 8 (February/March): 45–48.
Zogby International. 1997. ''Zogby Polls NJ and VA Right Again!'' Press release. Zogby International, Utica, NY.